

**Summary.** Two canonical measurements were extracted from single perspective scene views exemplifying minimal structures--scene measurements, as the interplanar distances between parallel planes, and the camera's position, with regards to the choice of reference planes. The minimal geometric traces required for this exercise are the vanishing line of a reference plane and a non-parallel direction's vanishing point. Affine information is generated without camera calibration or pose underpinnings. Collected outcomes are presented as 3D models.

**Necessary Geometries.** The vanishing line,  $\mathbf{l}$ , is the projection of the reference plane into the image as the line at infinity. The vanishing point,  $\mathbf{v}$ , is the image of the point at infinity in the reference direction. The reference plane can be intuitively considered the ground and the vanishing line the horizon.

**Projection Matrix.** The distance in the reference direction between two parallel planes with points  $\mathbf{x}$  and  $\mathbf{x}'$  in correspondence is the desired quantity to expose. To begin, a 3-dimensional coordinate system is staged, its x- and y- arms flush against and spanning the reference plane. The z-axis is a non-parallel, vertical reference direction. The vanishing points in the x-, y-, and z- directions, denoted  $\mathbf{v}_x$ ,  $\mathbf{v}_y$ , and  $\mathbf{v}$ , comprise the first three columns of  $\mathbb{P}$ , the  $3 \times 4$  projection matrix that projects a point  $\mathbf{X}$  in space to the image point  $\mathbf{x}$ ,

$$\mathbf{x} = \mathbb{P}\mathbf{X} = [\mathbf{p}_1 \mathbf{p}_2 \mathbf{p}_3 \mathbf{p}_4]\mathbf{X}$$

The fourth and final column of  $\mathbb{P}$  is the world coordinate system. As the coordinate frame has the x- and y- axes resting on the reference plane,  $\mathbf{v}_x$  and  $\mathbf{v}_y$  are two distinct points on the vanishing line. Columns 1, 2, and 4 of the projection matrix compose the three column homography of the reference plane to image. This homography must have rank 3, otherwise the map is degenerate. Consequently, the column of the origin of the coordinate system must not additionally lie on the vanishing line, else all three columns are not linearly independent. The final parameterization of projection matrix  $\mathbb{P}$  is

$$\mathbb{P} = [\mathbf{p}_1 \mathbf{p}_2 \mathbf{a} \hat{\mathbf{l}}]$$

with  $\hat{\mathbf{l}} = \mathbf{l} / \|\mathbf{l}\|$  and scale factor  $\mathbf{a}$ .

**Distance Measurements.** Retrieving interplanar measurements is independent of the first two generally under-determined columns of  $\mathbb{P}$ . Given  $\mathbf{v}$  and  $\mathbf{l}$ , the remaining unknown quantity is  $\mathbf{a}$ . An affine coordinate grid is assembled within the plane using the first, second, and fourth columns of  $\mathbb{P}$  to conduct coordinate measurements.

The distance between scene parallel planes specified as points  $\mathbf{X} = (X, Y, 0)^T$  and  $\mathbf{X}' = (X, Y, Z)^T$  is solved through

$$\mathbf{a}Z = -\|\mathbf{x} \times \mathbf{x}'\| / (\hat{\mathbf{l}} \cdot \mathbf{x}) \|\mathbf{v} \times \mathbf{x}'\|$$

Once  $\mathbf{a}$  is known, the metric value for  $Z$  is recovered through

$$Z = -\|\mathbf{x} \times \mathbf{x}'\| / (\mathbf{p}_4 \cdot \mathbf{x}) \|\mathbf{p}_3 \times \mathbf{x}'\|$$

Oppositely, a known (i.e. real world measured, when available) or estimated (i.e. under the circumstances of renaissance paintings or antiquated photographs) reference distance for  $Z$  to resolve affine ambiguity  $a$ .

**Algorithm.** Proceeds as described,

1. Two beginning and end points of sets of lines parallel to the x-, y-, and z-axes are selected.
2. The lines' intersections at their points at infinity situate the vanishing points.
3. The cross of the x- and y- directions' vanishing points sequence the vanishing line, of which the unit vector is retained.
4. The bottom and top points of a proposed reference distance are exacted.
5. The scalar quantity is resolved using the reference distance, its bottom and top points, the vertical vanishing point, and the unit vector of the vanishing line.
6. The top line of a desired distance is delineated through identifying two points which lie on it.
7. The vanishing point associated with the top line and the vanishing line is computed.
8. The bottom line of the desired distance is constrained to meet at the vanishing point and is decided after the remaining point nomination.
9. The point correspondence between the top and bottom lines of the desired distance is found.
10. The desired distance is accomplished using the corresponding points, the scalar quantity, the vertical vanishing point, and the unit vector of the vanishing line.
11. Repetition with alternate choices of reference plane recuperates the scalar quantity along each reference direction.

**Camera Location.** The completed projection matrix  $P$  additionally elicits camera location. The camera center,  $\mathbf{C} = (X_c, Y_c, Z_c, W_c)^T$ , is defined through Cramer's rule,

$$\begin{aligned} X_c &= -\det[\mathbf{p}_2 \ \mathbf{p}_3 \ \mathbf{p}_4], \\ Y_c &= \det[\mathbf{p}_1 \ \mathbf{p}_3 \ \mathbf{p}_4], \\ Z_c &= -\det[\mathbf{p}_1 \ \mathbf{p}_2 \ \mathbf{p}_4], \\ W_c &= \det[\mathbf{p}_1 \ \mathbf{p}_2 \ \mathbf{p}_3] \end{aligned}$$

**3D Modelling.** The second CAP6419 assignment was repurposed for affine and metric rectification to recover planar scene textures when able. GIMP-2.10's Projection tool facilitated surface reconstruction after dissatisfactory attempts toward precise metric plane rectification [Liebowitz, Criminisi, and Zisserman 1999].

The strategy attempted is--broadly--rectifying the vanishing points by the homography  $P = [1 \ 0 \ 0; 0 \ 1 \ 0; l_1 \ l_2 \ 1]$  involving the vanishing line. Rotating the first vanishing point by  $P$  such that it is aligned with the horizontal axis. The first affine transformation restores metric geometry up to an unknown aspect ratio is  $A_1 = [1 \ -\cot(\theta) \ 0; 0 \ 1 \ 0; 0 \ 0 \ 1]$ . The second affine transformation corrects the first using a lengths ratio in the vanishing points' directions,  $A_2 = [\mu \ 0 \ 0; 0 \ 1 \ 0; 0 \ 0 \ 1]$ .

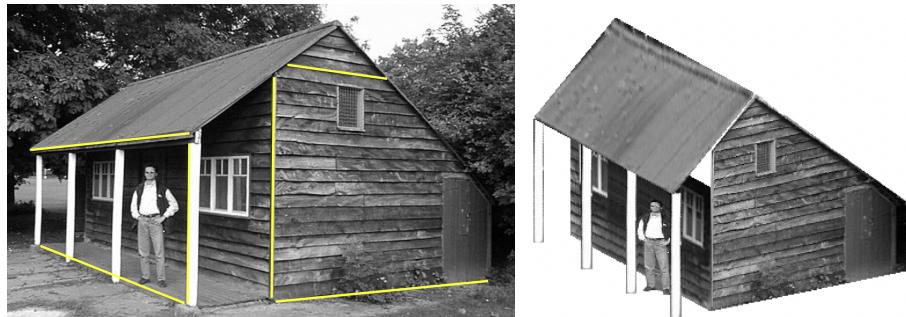
The correction  $\mu$  was not confidently recuperable during image investigations. While height was credibly produced in most cases, depths or widths along one of the parallel directions suffered noticeable discrepancies as to evince horizontal values dubious.

Modeling was executed through MATLAB's Surface plot functionality.

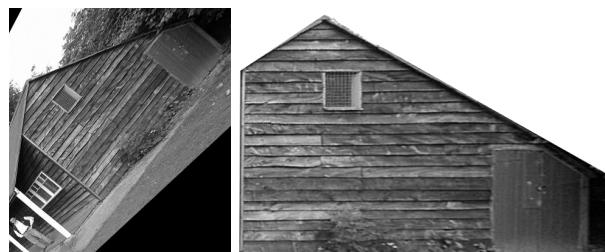
**Shed Example.** The reconstruction of the same single photograph of a shed that was examined in the paper is shown below [Criminisi, Reid, Zisserman]. The paper's provided reference height, along with reasoned width assumptions of the windows were established to output an approximate semblance.

The left face of the shed where the man is standing proved difficult from which to extract interplanar depths. The measurements expressed exorbitant deviations and instabilities. Eventual dimensions which presented visually "correct" were settled.

The computed camera height and relation explicates the structure's tapering away. Additionally, the man is roughly eye-level with and facing the purported camera's direction.



**Figure 1.a.** "shed.jpeg" with parallel lines highlighted (left) and collated 3D model (right).



**Figure 1.b.** Affine- and metric-rectified view (left) and planar texture of shed face (right).



**Figure 1.c.** Estimated camera frontal view and vantage.

**Painting Example (Part I).** An examination of the illustration of “Saint Jerome in His Study” proceeds. As with the previous shed image, it was similarly arduous to recover depth values. For several iterations, the lion was calculably far, far forward. A reconcilable placement was eventually settled.

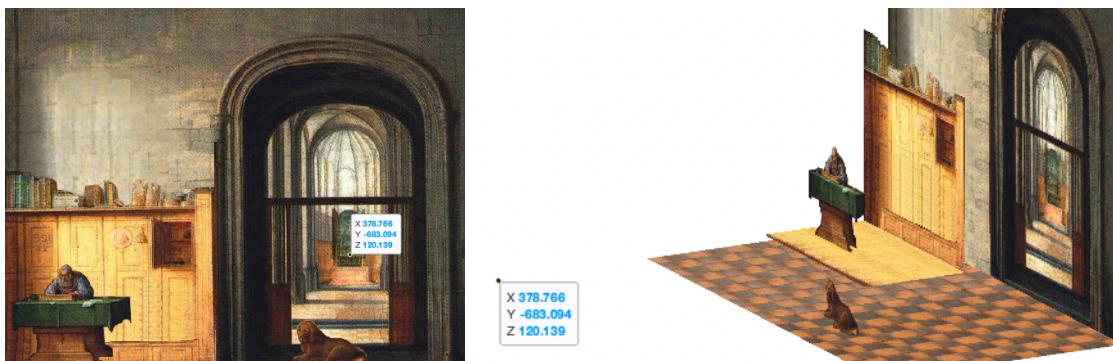
The camera location is suggested to be aligned with the doorway, which is cognitively in accordance with the hall’s distant, enduring visibility.



**Figure 2.a.** “painting.jpg” with parallel lines highlighted (left) and collated 3D model (right).



**Figure 2.b.** “painting.jpg” (left) and projectively restituted floor texture (right).



**Figure 2.c.** Estimated camera frontal view and vantage.

**Painting Example (Part II).** Below lies a reattempted composition with two additional textures to further investigate the metrology in the gazing direction.

An emerged possibility is that the camera is in the crux of the scene and the lion and front edge of the fireplace is located behind the camera. An echoing claim is that the lion appears to “float” atop the tiles. The floor and ceiling do not linearly widen nor fan out as they nearer the viewer--their lengths suspend in the midst of the painting.

The camera is positioned slightly left of the doorway, resonating the doorway view’s veering bias to the right.



**Figure 3.a.** “painting.jpg” (left) and collated 3D model (right).



**Figure 3.b.** Zooms of the tiles unveil that drawn lengths are not perceptibly incrementing.



**Figure 3.c.** Estimated camera frontal view and vantage.

**Stereogram Example.** The below is a cursory yet sincere endeavor to address a problem pitched by Amy Giroux during the Fall 2021 wave of COP4934 senior design project proposals. It was viscerally one of my coveted topics but outlying in my choice of assignments as I had anticipated it to be quite the undertaking.

St. Augustine National Cemetery holds three pyramid tombs that were erected in 1842. Beside are headstones placeholdering unidentified soldiers. Twenty men from the 7th Regiment of the New Hampshire Volunteers who were stationed in St. Augustine during the Civil War lost their lives.

The wooden headboards of the deceased decomposed before marble headstones were erected. From a stereo image (i.e. half of a stereogram) of St. Francis Barracks in 1865, the distances between the leftmost pyramid and the headboards in the image are to be ascertained to align grave ownership.

Average tombstone dimensions and separation between plots was assumed. The result is visually acceptable and depicts approximate proportionality upon inspection against a modern photo of the cemetery. The camera location is reasonable and supports the projectivity of the farther graves, pyramids, and pillar “shrinking” and the posterior objects’ obfuscation.



**Figure 4.a.** Stereogram (left) and the three pyramids and marble headstones (right) from Giroux's presentation.



**Figure 4.b.** Names that were legible and their respective graves.



**Figure 4.c.** “gravesite.png” with parallel lines highlighted (left) and collated 3D model (right).



**Figure 4.d.** Comparison of modern image (left) with model suggests, at the least, metric fidelity along the vertical axis.



**Figure 4.e.** Estimated camera frontal view and vantage.

**Discussion.** After studies of several single scene images with minimal structural information, there is a degree of reconstructability from applying the described methodology. The horizontal and depth interplanar measurements were more difficult to ascertain than those along the vertical direction.

Measurements from single scene images are most easily deducible when one real world reference distance in each direction is available. The 3D reconstruction of the gravesite from the half stereogram would be augmented from knowing a groundtruth measurement between the pillar and pyramids, as the wooden headboards are irretrievable. Thus, a practical application of single image metrology is recovering relational distances after some scene items have been removed.

**Further Study.** Without time limitations, experiential application of the following would be pursued.

**Robust Computation of Vanishing Points.** As model fidelity relies on the precision of the vanishing points, an enhanced strategy for obtaining the “best fit” intersection of  $n$  parallel lines is sought.

The  $3 \times 3$  second moment symmetric matrix  $\mathbf{M}$  of competing line candidates is represented

$$\mathbf{M} = \text{sum}[x_i x_i \ x_i y_i \ x_i z_i; \ x_i y_i \ y_i y_i \ y_i z_i; \ x_i z_i \ y_i z_i \ z_i z_i]$$

where  $i$  is the  $i$ -th homogeneous vector and onto which the Jacobi-method eigendecomposition is performed. The eigenvector associated with the smallest eigenvalue is the vanishing point. John Burkardt’s version can be employed for this effort to refine the model at a later date.

**Internal Parameters.** The image of the absolute conic, an esoteric entity on the plane at infinity in the third dimension, offers a functional object for reasoning orthogonality and is related to camera calibration,  $\omega = K^T K^{-1}$ . A typical real world application, such as an image of a structure, gives three orthogonal directions and constraints on  $\omega$ .

The orthogonality equations for a natural camera are equivalent to a single construction, the triangle with three orthogonal vanishing points as its vertices and the principal point as its orthocenter. The coefficient matrix  $A$

$$A^T = [ \begin{array}{ccccc} u_1 v_1 & u_1 w_1 & v_1 w_1 & 0 & 1; \\ u_1 v_2 + u_2 v_1 & u_1 w_2 + u_2 w_1 & v_1 w_2 + v_2 w_1 & 1 & 0; \\ u_2 v_2 & u_2 w_2 & v_2 w_2 & 0 & -1; \\ u_1 v_3 + u_3 v_1 & u_1 w_3 + u_3 w_1 & v_1 w_3 + v_3 w_1 & 0 & 0; \\ u_2 v_3 + u_3 v_2 & u_2 w_3 + u_3 w_2 & v_2 w_3 + v_3 w_2 & 0 & 0; \\ u_3 v_3 & u_3 w_3 & v_3 w_3 & 0 & 0 ]$$

facilitates calculation of  $\omega$ , as a null vector,  $A\omega = 0$ , and thus the assemblage of the symmetric matrix  $\omega$ . The Cholesky decomposition, a factorization that decomposes a symmetric matrix into the product of a lower and upper triangular matrix, of its inverse provides the upper triangular camera calibration matrix  $K$ , the equivalence to knowing the camera internal parameters.

A simplified camera model with zero skew and unit aspect ratio natural camera is

$$K = [f \mathbf{0} \mathbf{u}_0; 0 f \mathbf{v}_0; 0 0 1]$$

### **References.**

- R. Collins, <https://www.cs.cmu.edu/~ph/869/www/notes/vanishing.txt>.
- A. Criminisi, I. Reid and A. Zisserman, Single View Metrology.
- A. Giroux, A Solution for Stereogram Measurements: CV, ML, or something else?
- D. Liebowitz, A. Criminisi and A. Zisserman, Creating Architectural Models from Images.
- J. Burkardt , <https://people.sc.fsu.edu/~jburkardt/>.